



Article

Spatio-Temporal Knowledge Graph Based Forest Fire Prediction with Multi Source Heterogeneous Data

Xingtong Ge ^{1,2} , Yi Yang ³, Ling Peng ^{1,2,*}, Luanjie Chen ^{1,2}, Weichao Li ¹, Wenyue Zhang ^{1,2} and Jiahui Chen ^{1,2}

¹ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; gexingtong21@mails.ucas.ac.cn (X.G.); chenluanjie20@mails.ucas.ac.cn (L.C.); liweichao@rsai.tech (W.L.); zhangwenyue@rsai.tech (W.Z.); chenjiahui21@mails.ucas.ac.cn (J.C.)

² College of Resources and Environment (CRE), University of Chinese Academy of Sciences, Beijing 100190, China

³ Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China; yangyi@ia.ac.cn

* Correspondence: pengling@aircas.ac.cn

Abstract: Forest fires have frequently occurred and caused great harm to people's lives. Many researchers use machine learning techniques to predict forest fires by considering spatio-temporal data features. However, it is difficult to efficiently obtain the features from large-scale, multi-source, heterogeneous data. There is a lack of a method that can effectively extract features required by machine learning-based forest fire predictions from multi-source spatio-temporal data. This paper proposes a forest fire prediction method that integrates spatio-temporal knowledge graphs and machine learning models. This method can fuse multi-source heterogeneous spatio-temporal forest fire data by constructing a forest fire semantic ontology and a knowledge graph-based spatio-temporal framework. This paper defines the domain expertise of forest fire analysis as the semantic rules of the knowledge graph. This paper proposes a rule-based reasoning method to obtain the corresponding data for the specific machine learning-based forest fire prediction methods, which are dedicated to tackling the problem with real-time prediction scenarios. This paper performs experiments regarding forest fire predictions based on real-world data in the experimental areas Xichang and Yanyuan in Sichuan province. The results show that the proposed method is beneficial for the fusion of multi-source spatio-temporal data and highly improves the prediction performance in real forest fire prediction scenarios.

Keywords: forest fire prediction; knowledge graph; spatio-temporal data; machine learning; multi-source heterogeneous data fusion



Citation: Ge, X.; Yang, Y.; Peng, L.; Chen, L.; Li, W.; Zhang, W.; Chen, J. Spatio-Temporal Knowledge Graph Based Forest Fire Prediction with Multi Source Heterogeneous Data. *Remote Sens.* **2022**, *14*, 3496. <https://doi.org/10.3390/rs14143496>

Academic Editor: Luke Wallace

Received: 29 April 2022

Accepted: 19 July 2022

Published: 21 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Forest fires are one of the major interference factors in forest ecosystems, which affect biodiversity, species composition, and ecosystem structure [1]. They threaten forest resources and the safety of lives and property [2]. Thus, the prediction of forest fires plays an important role in disaster risk analysis. This paper proposes a spatio-temporal knowledge graph framework that can improve machine learning-based forest fire prediction methods. The proposed framework can deeply extract and analyze features of a large amount of historical spatio-temporal data, particularly for the rapid changes in fire information and incomplete fire data.

Commonly used approaches include fire indices and mechanism models for forest fire predictions that consider the closely related occurrence factors of forest fires. Salavatiet et al. [3] conducted research in the city of Sanandaj, located in the west of Iran. In this study, fire risk potential is assessed using Weights of Evidence (WoE) and Statistical Index (SI) models. Chen et al. [4] considered precipitation as an important factor affecting the probability of forest fire occurrence. In their work, they present a method for better representing the effect of precipitation on predicting forest fires. Ge et al. [5] proposed a comprehensive index

of forest fire drivers for forest fire predictions using a hierarchical analysis considering topographic, vegetation, meteorological, and human activity factors.

Recently, quite a few studies have shown that machine learning-based methods can discover more complex data patterns for forest fires than the traditional mechanism- or statistic-based models. Jaafari et al. [6] employed WoE Bayesian modeling to investigate the spatial relations between historical fire events in the Chaharmahal-Bakhtiari Province of Iran. Sakr et al. [7] presented a forest fire risk prediction algorithm based on Support Vector Machines (SVM). To predict forest fires, Binh Thai Pham et al. [8] compared the ability of Bayes Network (BN), Naive Bayes (NB), Decision Tree (DT), and Multivariate Logistic Regression (MLP) to map fire susceptibility in Phu Mat National Park, Ninh An Province, Vietnam. Ma et al. [9] built a forest fire probability model based on the Logistic model and the Random Forest (RF) model with the forest thermal anomaly data monitored by satellites from 2010 to 2017. The model analyzes the driving factors of forest fires in the Shanxi Province of China from four aspects: meteorology, topography, vegetation, and human activities. Singh et al. [10] utilized the RF approach for assessing the impacts of the climatic and anthropogenic factors for influencing fire occurrence probability and mapping the spatial distribution of fire risk. Prapas et al. [11] considered daily fire danger prediction as a machine learning task using historical Earth observation data from the last decade to predict the next-day's fire danger. Their Deep Learning-based method provides nationwide daily fire danger maps, with a much higher spatial resolution than the existing operational solutions. The previous research provides a good foundation for forest fire prediction using machine learning methods.

Remote sensing technology can capture a large amount of dynamic information in large observation ranges. The forest fire prediction studies make good use of remote sensing techniques. The MODIS (Moderate-resolution Imaging Spectroradiometer) Fire and Thermal Anomalies data and the VIIRS (Visible Infrared Imaging Radiometer Suite) Fire data show active fire detections and thermal anomalies [12]. Cui et al. [13] used the Terrestrial Water Storage Change (TWSC) generated by Gravity Recovery and Climate Experiment (GRACE) data to analyze the influence of climate change on forest fires in the region between 2003 and 2016. Piralilou et al. [14] evaluated the effects of coarse (Landsat 8 data and SRTM data) and medium (Sentinel-2 data and ALOS data) resolution spatial data on wildfire susceptibility prediction using models based on RF and SVM.

There are many data sources for forest fire predictions. There are barriers for data such as time, space, type, resolution, and coordinate systems. There is a difficult critical problem that needs to fuse the large-scale multi-source heterogeneous data related to forest fires for deeply discovering data patterns. In particular, the spatio-temporal features and attribute features extracted from the data need to be semantically defined. The semantics of the fused data can be beneficial to the spatio-temporal query of data, as shown in Table 1. In addition, the proposed forest fire prediction method requires expertise in fire prediction, which includes forest fire occurrence factor determination, forest fire prediction model optimization strategies, and forest fire prediction model accuracy evaluation. The existing forest fire prediction methods rarely apply semantic descriptions for fire prediction expertise. Therefore, we cannot directly match model parameters or hyper-parameters that are most suited for real-world forest fire prediction scenarios. To address the above issues, we focus on forest fire expertise modeling and rule-based semantic reasoning for forest fire predictions.

In this paper, we propose a forest fire prediction method by considering meteorological, topographic, vegetation, and human activity factors that are closely related to forest fire occurrence. This method builds a knowledge graph for Forest Fire Prediction (KGFFP) to fuse heterogeneous forest fire data from multiple sources. This method differs from forest fire predictions that are only based on machine learning methods. It formalizes the expertise in the field of forest fire as semantic rules and obtains predicting data from KGFFP through a rule-based reasoning method. The predicting data fit specific machine learning models that work as cores of forest fire predictions. The proposed method builds on traditional

machine learning-based forest fire prediction. It can fuse multi-source heterogeneous spatio-temporal data and formalize domain expertise. The proposed method helps to improve the efficiency of predictive analysis based on semantic inference rules. Figure 1 shows the comparison between the proposed method and the forest fire predictions based on machine learning methods.

Table 1. Comparison between the proposed method and the commonly used data fusion methods for forest fire prediction.

Commonly Used Methods	The Method Proposed in This Paper
Through data processing, the time, space, type, resolution, and coordinate systems of data are unified.	From the semantic point of view, the temporal, spatial, and attribute features of the data are fused.
The spatio-temporal data is stored in a relational database and needs to be added, deleted, queried, and changed based on the graph database.	The spatio-temporal data is stored in the graph database and needs to be added, deleted, queried, and changed based on the graph database.
Multi-table joins are required for in-depth searches in relational databases, resulting in low query efficiency.	Based on the graph database, in-depth queries can be made from a limited space-time region. The efficiency of querying various geographic entities and the relationship between geographic entities in forest fire prediction scenarios is high.

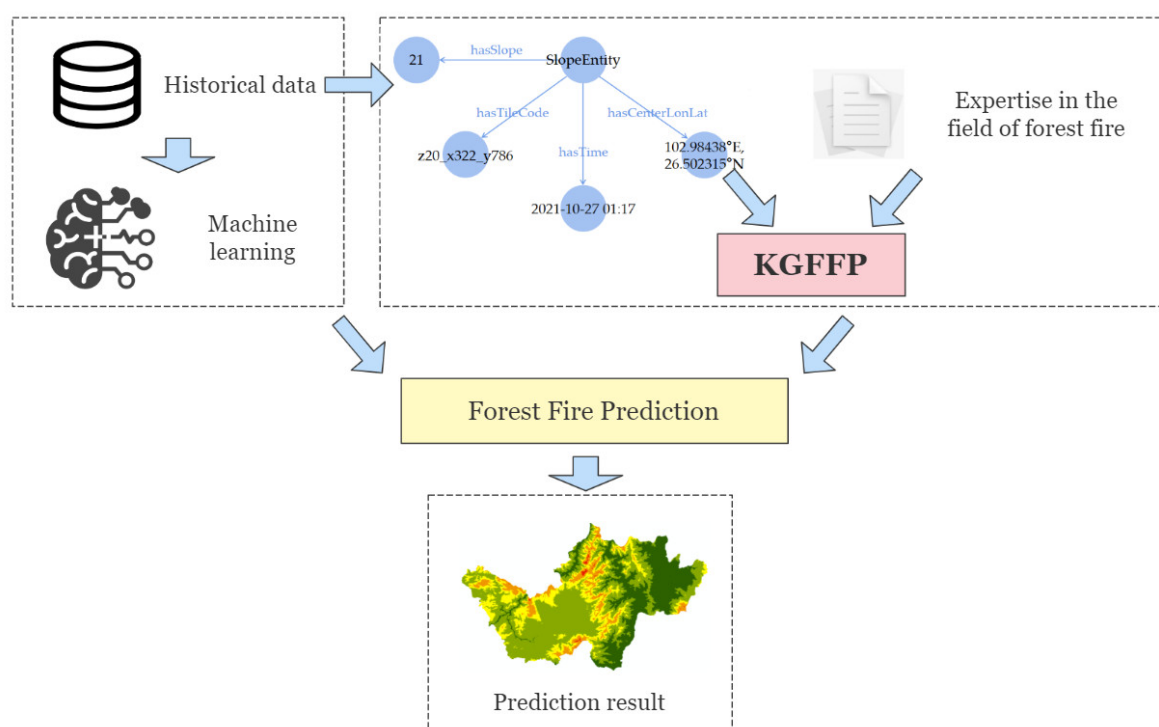


Figure 1. The comparison between this method and the forest fire prediction based on machine learning method.

This paper has the following contributions: (1) we build a knowledge graph-based multi-source heterogeneous data fusion method; (2) we define ontologies for modeling domain expertise of the forest fire risk analysis as semantic rules; (3) we propose a rule-based reasoning method for obtaining the corresponding predicting data required by the machine learning-based forest fire prediction methods from KG according to the specific situations; (4) we show experiments for demonstrating the benefits of the proposed method in the aspects of multi-source heterogeneous spatio-temporal data fusion and machine learning-based forest fire predictions.

2. Knowledge Graph-Based Forest Fire Prediction System (KGFFP)

A knowledge graph is a semantic modeling method for data representation and modeling through defining entities, concepts, and semantic relations. The rule-based reasoning over knowledge graphs can analyze the entities and relations over knowledge graphs by the specified semantic rules.

The architecture of the proposed forest fire prediction method consists of three basic parts: KGFFP, machine learning-based forest fire prediction algorithms, and a controller which controls rule inference and program flow. Sections 2.1 and 2.2 mainly introduce the construction of KGFFP. The machine learning-based forest fire prediction algorithms in this paper directly use the framework proposed by other existing papers. The controller focuses on the rule-based reasoning for the semantic rules. Section 2.3 shows the implementation of the proposed method.

2.1. Method Architecture

2.1.1. Conceptual Model for Forest Fire Prediction

The occurrence of forest fires is closely related to meteorological, topographical, vegetation, and human activity factors [15,16]. In this paper, we build a conceptual model for forest fire prediction considering the above factors. It provides mechanism information for forest fire prediction, including data that are used in KGFFP, as shown in Figure 2. The data type will be a difference between the changes in the research objects and the actual situations.

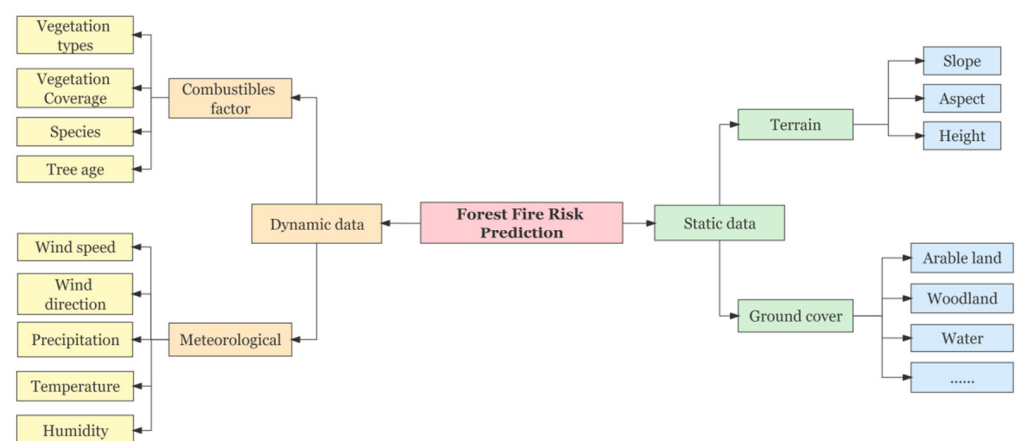


Figure 2. Conceptual data model for KGFFP Architecture.

2.1.2. Mapping between Forest Fire Prediction and KGFFP Architecture

We propose the KGFFP architecture, as shown in Figure 3, to support the fusion of multi-source heterogeneous spatio-temporal data related to forest fires. The architecture can match parameters and data for a specific machine learning-based forest fire prediction method.

The KGFFP fuses the spatio-temporal data of forest fire by representing temporal features, spatial features, and attribute features of spatio-temporal data. Therefore, time ontology and space ontology need to be built for the conceptual layer of the knowledge graph. The attribute features of spatio-temporal data strongly depend on the data type (e.g., aspect is one of the important features of terrain data) that have specific hierarchical relations for forest fire predictions (e.g., topographic data include slope, aspect, and elevation). The conceptual model for forest fire predictions is defined as a part of the conceptual layer of the knowledge graph. The unstructured, structured, and semi-structured spatio-temporal data for forest fire predictions can be modeled as a part of the instance layer of the knowledge graph. The temporal features, spatial features, and attribute features of the spatio-temporal data are represented based on the conceptual layer of the knowledge graph. The concepts and actions used by the machine learning-based forest fire predictions

are defined as an ontology. According to the concept definition in ontology, the domain knowledge is formalized into the semantic reasoning rules of the knowledge graph.

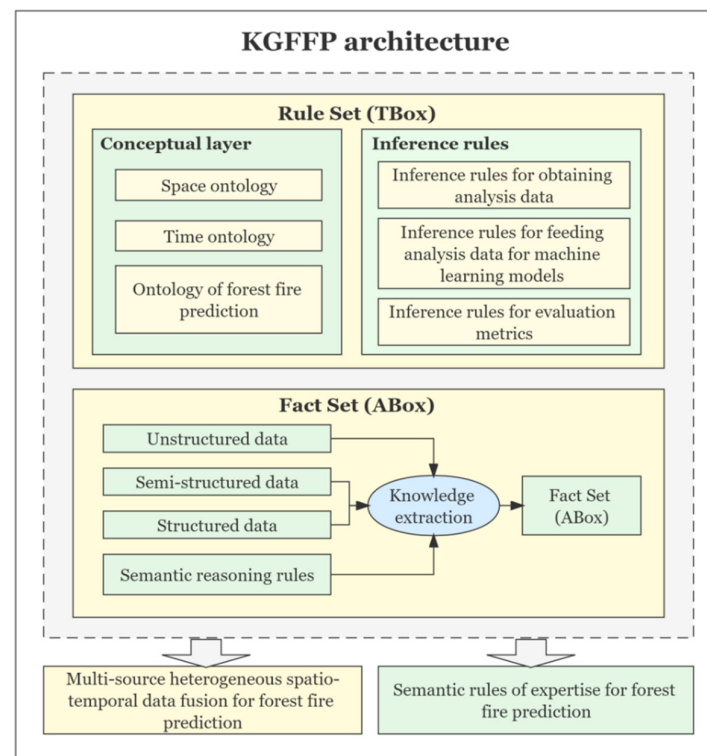


Figure 3. KGFFP architecture.

The KGFFP consists of the rule set (TBox) and the fact set (ABox). The TBox consists of the conceptual layer and the inference rules. The conceptual layer is the semantic basis for representing the hierarchical relationship of various factors regarding forest fire predictions. The inference rules are the logic basis for multi-source spatio-temporal data that supports semantic reasoning. The ABox constitutes the instance layer of the knowledge graph, which contains instances corresponding to various concepts in the concept layer. The ABox and the TBox together constitute the reasoning mechanism of the KGFFP.

2.2. Construction of Forest Fire Prediction Knowledge Graph

This section introduces the design of the concept layer, instance layer, and inference rules of KGFFP. We use Ontology Web Language (OWL) as the knowledge representation language as it is a widely used expressive language for knowledge graphs [17].

2.2.1. Design of Conceptual Layer

The conceptual layer of KGFFP defines the logical semantic relationship among multi-source spatio-temporal data for forest fire predictions. It contains semantic concepts and their interrelationships. It ensures the consistency of semantic concepts inherent in multi-source spatio-temporal data related to forest fire predictions. The concept layer of KGFFP includes time ontology, space ontology, an ontology for forest fire predictions, and machine learning model concept ontology.

- Time ontology and space ontology

We used the time ontology of Semantic Web Rule Language (SWRL) [18] to represent the common time concept of KGFFP. SWRL provides a uniform specification of the semantic representation of time. It ensures the comparability and computability of the temporal information of entities [5].

We built a space ontology based on the extensions of Geographic Query Language (GeoSPARQL) [19].

- Ontology for forest fire prediction

To define the conceptual model of forest fire predictions as forest fire prediction ontology, we propose a tree-like taxonomy for the conceptual partitioning of multi-source spatio-temporal data [5]. We constructed forest fire prediction ontology according to the KGFFP architecture, as in Figure 4.

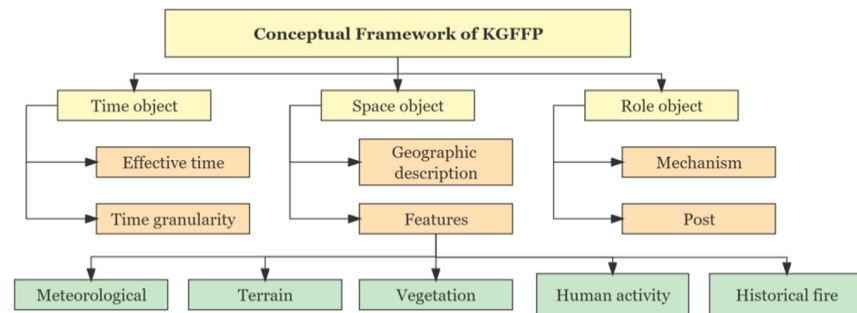


Figure 4. Tree-like taxonomy for conceptual partitioning of multi-source spatio-temporal data.

- Machine learning model concept ontology

To improve the task of the machine learning-based forest fire predictions, we construct the concept ontology of the machine learning model. It provides a conceptual framework for the optimization and prediction of the machine learning model, as shown in Figure 5. The architecture is applicable to different supervised machine learning models, including but not limited to RF and Deep Forest (DF).

Machine Learning Model Concept Ontology

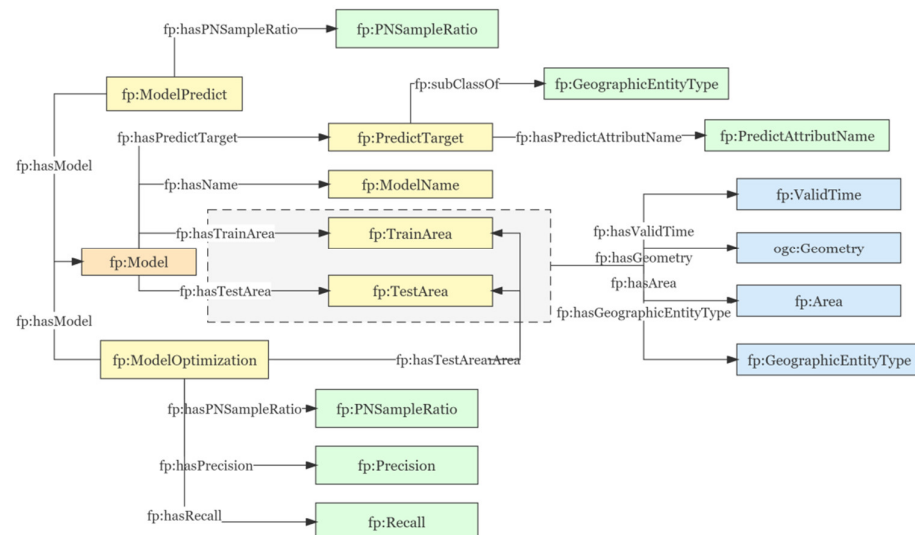


Figure 5. Framework of machine learning model conceptual ontology.

In the model, *fp* is short for fire predict, a prefix for named entities in KGFFP. *fp : ModelOptimization* is a class that represents the optimization stage of the machine learning model. Both *fp : Precision* and *fp : Recall* denote evaluation metrics, which are objects associated with *fp : ModelOptimization*. *fp : ModelPredict* is a class that represents the prediction stage of the machine learning model.

fp : Model is the object of *fp : ModelOptimization* and *fp : ModelPredict*, which represents a machine learning model. Its objects include the model name, prediction target,

training area, and testing area. Both training and testing regions are associated with their temporal phase, spatial geometry, regions, and properties of the analytical data to which the model applies. The controller instantiates the above concepts when performing training or prediction.

2.2.2. Design of Instance Layer

- Instantiation of Multi-source spatio-temporal data

Multi-source heterogeneous data in the field of forest fire prevention often have different temporal characteristics. In this paper, multi-source heterogeneous forest fire data are classified into dynamic data and static data according to the update frequency and accessibility, as shown in Figure 2. Dynamic data are indexed according to the hierarchical relationship of year, month, day, hour, minute, and second, while static data usually have only one phase. Data with the same properties can be converted between dynamic and static data according to the changes in update frequency and accessibility.

The data classification can facilitate data modeling with different time resolutions and changes in time resolutions. We can query the multi-source spatio-temporal data by considering the time resolution. Figure 6 shows a multi-source spatio-temporal data query task. The task objective is to query the temperature, vegetation coverage, and slope in Xichang City at 9:00 on 30 March 2020. First, it determines whether the three types of data are dynamic data or static data. Temperature and vegetation coverage are treated as dynamic data, and the slope is classified as static data. Then, it queries for finding a slope with a unique phase. If it exists, the slope acquisition is successful; if it does not exist, the slope acquisition fails. Finally, it queries the temporal resolutions for temperature and vegetation coverage, respectively. When the time resolution of the temperature is an hour, the temperature on 30 March 2020 at 9:00 are obtained. When the temporal resolution of vegetation coverage is month, and the vegetation coverage in March 2020 are obtained.

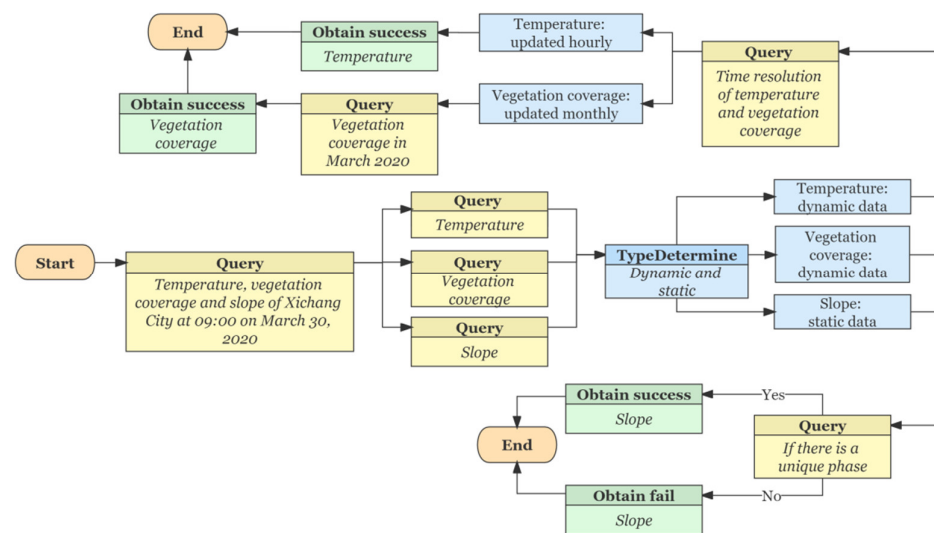


Figure 6. Matching of multi-source data in the time dimension.

- Instantiation of machine learning-based forest fire prediction model

When the controller performs the optimization and prediction of the machine learning-based forest fire prediction method, instances of the concept ontology of the machine learning model are built, as shown in Figure 7.

In the model optimization phase, first, it selects a machine learning model, instantiates $fp : ModelOptimization$, and instantiates $fp : Model$; next, it instantiates $fp : ModelName$, $fp : PredictTarget$, $fp : TrainArea$ and $fp : TestArea$; then, it determines $fp : PNSampleRatio$ (the ratio of positive samples and negative samples) and optimizes the model; finally, it instantiates $fp : Precision$ and $fp : Recall$ according to the test accuracy.

In the model prediction stage, first, it selects a machine learning model, instantiates $fp : ModelPrediction$, and instantiates $fp : Model$; next, it instantiates $fp : ModelName$ and $fp : PredictTarget$; finally, it determines $fp : PNSampleRatio$ for model predictions.

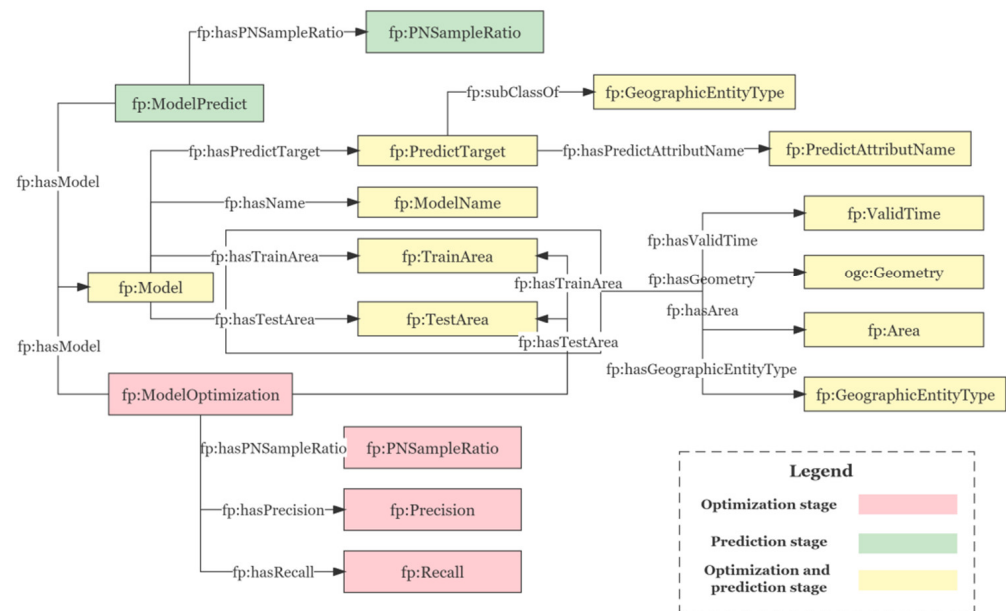


Figure 7. Machine learning-based forest fire prediction model instantiation.

2.2.3. Design of Inference Rules of Forest Fire Predictions

This section introduces methods for representing forest fire prediction expertise as semantic inference rules in knowledge graph.

There is a lot of expertise used in forest fire predictions. Domain knowledge includes data types related to forest fire prediction, data preprocessing and fusion methods, and prediction strategies of machine learning models. Usually, it is necessary to determine whether the existing situation satisfies the predetermined triggering conditions. If yes, the system will take corresponding response actions. Response actions will, in turn, generate new trigger events. For example, when a historical fire point is detected by the controller, it requires the land cover types corresponding to the spatio-temporal range where the fire point is located. If the land cover type is a built-up area, the fire point is rarely possible recognized as a forest fire. Therefore, we formalize domain expertise as a set of rules, each of which is described in the form of *RuleObject*. A schedule object defines the events that need the triggers and the actions. Class is entitled by the event name and action name. Classes are related to each other through defined data properties or object properties. Figure 8 shows the syntax of *RuleObject*.

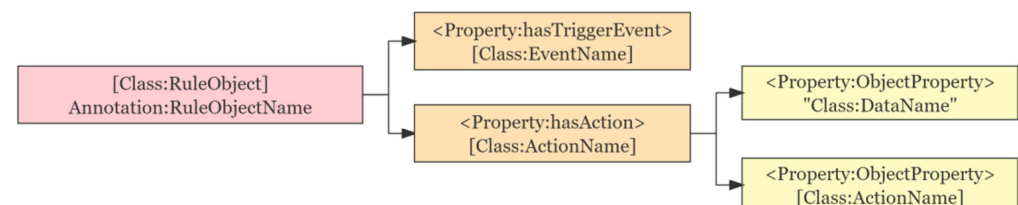


Figure 8. Syntax of rule definition.

We extracted actions, sequences between actions, and trigger relations between actions from expertise. The controller defines the trigger events and actions of the *RuleObject* based on the above information. The principles of rule definition include (1) Expressiveness: Representing the association between different actions and different rule objects; (2) Reusability: Designing rules using defined data attributes and object attributes as many as possible.

Figure 9 shows the rules of data extraction for machine learning models from multi-source spatio-temporal data (partial).

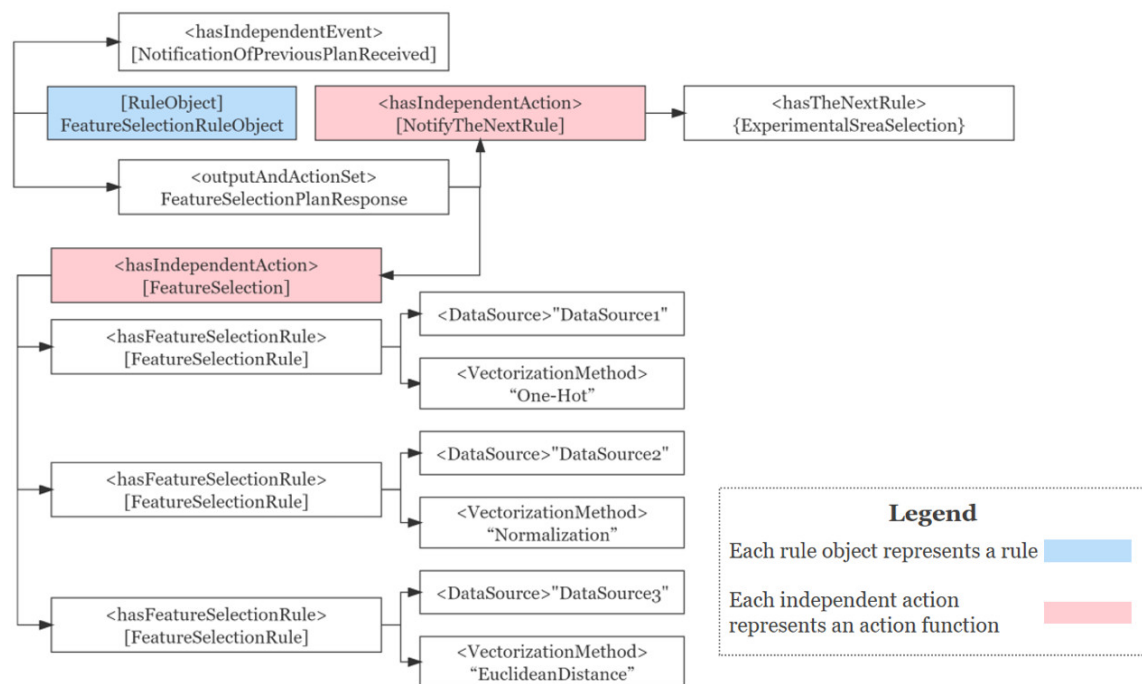


Figure 9. The rule set for extracting data for machine learning models from multi-source spatio-temporal data (partial).

2.3. Implementation of Spatio-Temporal Knowledge Graph for Forest Fire Prediction

This section shows the implementation of the proposed forest fire prediction method. Figure 10 shows the architecture.

The architecture consists of four basic layers. In the data resource layer, large-scale heterogeneous spatio-temporal data are collected from different data sources. In the knowledge extraction layer, we provide different automatic knowledge extraction methods respectively for the data with different structures. The structured data, semi-structured data, and unstructured data are automatically converted into GeoJSON format. In the knowledge storage layer, the conceptual framework of forest fire prediction ontology is constructed by the Protégé tool according to relations of data for forest fire prediction in the conceptual layer. The ontology data are stored by GraphDB. Based on the forest fire prediction ontology, the forest fire data are converted into triples for building the instance layer. The instance layer is stored by the Key-Value databases. In the analysis service layer, with the help of the KGFFP, it provides services of forest fire prediction by using the spatio-temporal semantic query technology.

2.3.1. Data Resource Layer

This layer is composed of multi-source heterogeneous raw data in the field of forest fire prediction. Forest fire prediction requires multi-source heterogeneous spatio-temporal data whose structures are different. Specifically, meteorology (temperature, humidity, wind direction, etc.) has structured data; fire point distribution and land cover type have semi-structured data; and professional knowledge (text) has unstructured data. The topographic data are static data with the lowest updating frequency. Vegetation data and vegetation coverage data need to be updated according to seasonal changes. The update frequency of land cover data is low. The shorter the time difference between the update time and the forecast time, the closer the land cover data are to the actual data. Meteorological data are

usually updated frequently, which can influence the time interval of forest fire prediction. Besides, the update frequencies of other data are low.

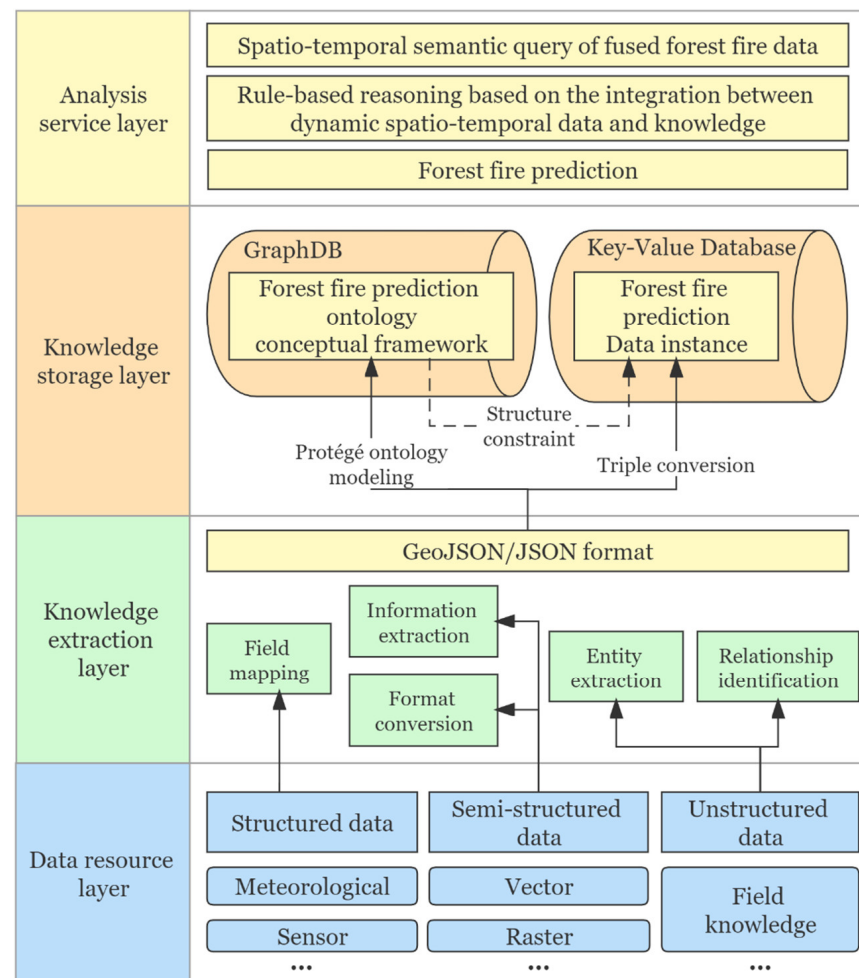


Figure 10. Architecture of forest fire prediction method based on spatio-temporal knowledge graph.

It is necessary to collect data according to the characteristics and update the frequency of the data in order to provide accurate and stable data resources for KGFFP.

2.3.2. Knowledge Extraction Layer

We used Protégé to design ontology for the conceptual layer, including time and space ontology. New concepts can be created including the hierarchical relations of classes, object attributes, and the data attributes of classes. The constructed ontology is stored as an RDF file.

We designed different triple transformation methods for different types of multi-source heterogeneous spatio-temporal data for forest fire predictions. It needed to be transformed to a unified coordinate system.

The commonly used vector data format in geographic information can be converted into a GeoJSON format using GDAL library. The time, space, and attribute information contained in GeoJSON can be converted into triples using the Arcpy [20] or GDAL library. It can respectively convert data from a raster gray value to an attribute in the vector data, and from vector format to GeoJSON format. Certain original data of discrete point distribution, e.g., weather station data are inconvenient for comparison with the distribution patterns of other spatial phenomena. Therefore, it uses an appropriate spatial interpolation model to generate raster-type interpolation results according to the distribution of point data. Then, the results are converted into GeoJSON format.

We built a converter that transforms Geometry in GeoJSON into a triple predicate and transforms Geometry values into objects that conform to the GeoSPARQL format specification; the keys of properties in GeoJSON correspond to the predicate, and the values correspond to the object of the predicate.

2.3.3. Knowledge Storage Layer

We used GraphDB to store knowledge. GraphDB is a highly efficient, robust, and scalable RDF database that can perform semantic reasoning at scale for massive loads, queries, and reasoning in real time.

2.3.4. Analysis Service Layer

We built the spatio-temporal semantic reasoning rules (*RuleObject*) according to the expertise and prediction algorithms. The rules are stored as triples. We used *SPARQL* for the spatio-temporal semantic query. In this way, it can realize the logical reasoning in *ActionObject* by the quantitative calculations of the query results.

3. Forest Fire Prediction Experiment Using KGFFP

This section presents experiments of KGFFP-based forest fire predictions through case study. These investigations show that the proposed method is beneficial for multi-source spatio-temporal data fusion, expertise formalization, and forest fire prediction.

3.1. Area of Experiment

We selected Xichang City and Yanyuan County in Liangshan Yi Autonomous Prefecture of China's Sichuan Province for a dynamic prediction experiment of forest fire disasters.

Xichang City is the capital of Liangshan Yi Autonomous Prefecture. It is located between $101^{\circ}46'$ – $102^{\circ}25'$ east longitude and $27^{\circ}32'$ – $28^{\circ}10'$ north latitude, with an area of 2882.9 square kilometers. Xichang City has high vegetation coverage and a dry climate, which is prone to forest fires. Historically, there have been forest fires in Xichang, which have caused casualties of rescuers. Therefore, we used Xichang City as the research area. Yanyuan County is adjacent to Xichang City. Yanyuan County is located between $100^{\circ}42'09''$ – $102^{\circ}03'44''$ east longitude and $27^{\circ}06'31''$ – $28^{\circ}16'31''$ north latitude, with a total area of 8398.6 square kilometers. There is rich vegetation and many shrubs in the area. The forest fire disasters are easily caused by inducing factors such as low rainfall. Therefore, for Yanyuan country, effective forest fire predictions can significantly reduce the damage caused by forest fire. Table 2 shows the sources and details of the experimental data.

Table 2. Sources and details of experimental data.

Data	Sources	Spatial Resolution	Temporal Resolution	Satellite Sensor
Fire	Fire Map–NASA [12]	About $0.375\text{ km} \times 0.375\text{ km}$; about $1\text{ km} \times 1\text{ km}$	Mid-latitudes will experience 3–4 looks a day	VIIRS S-NPP; VIIRS NOAA-20; MODIS/Aqua; MODIS/Terra
Meteorological	ERA5-Land hourly data from 1950 to present [21]	$0.1^{\circ} \times 0.1^{\circ}$; Native resolution is 9 km	Update hourly	None
Terrain	Shuttle Radar Topography Mission DEM [22]	$30\text{ m} \times 30\text{ m}$	Acquired 11–22 February 2000	STS Endeavour OV-105
Land cover	Esri_2020_Land_Cover_V2 ImageServer [23]	$10\text{ m} \times 10\text{ m}$	Released in 2020	Sentinel-2L2A/B
Vegetation	“One Map” of Forest Inspection and Forest Resource Management in 2020 [24]	$2\text{ km} \times 2\text{ km}$	Released in 2020	None
Normalized Difference Vegetation Index (NDVI)	China Quarterly Vegetation Index (NDVI) Spatial Distribution Dataset [25]	$1\text{ km} \times 1\text{ km}$	Updated quarterly	SPOT; VEGETATION; MODIS

3.2. Predictive Model

We applied RF in the experiment that respectively builds decision trees for samples that are extracted by using the bootstrap resampling method. According to The Law of Large Numbers, RF has less overfitting [26–28]. The dependent variable of RF prediction represents whether the forest fire occurs. The value of 1 represents forest fire occurrence, while value of 0 represents no occurrence. Therefore, forest fire predictions with RF can be treated as a binary classification task.

We applied DF as one of the core machine learning models of the proposed forest fire prediction. In recent years, in the field of computer vision, DF first uses convolutional neural networks (CNNs) to extract deep features [29,30] and then uses RF as a classifier [31,32].

3.3. Construction of KGFFP

We built the concept layer of KGFFP using Protégé in the case study, including time ontology, space ontology, forest fire prediction hierarchical ontology, and machine learning model concept ontology. It provided a conceptual framework for modeling the time, space, and attribute characteristics of data.

In addition, we constructed the instance layer of the spatio-temporal knowledge graph. March and April are the periods that are prone to forest fires. In March 2019 and March 2020, Xichang City had a large-scale forest fire for two consecutive years. We selected the data from March and April 2015–2020 as the research object because there is no access to the data for these two years, i.e., the data from 2021 and 2022 are not used as research objects. We collected multi-source heterogeneous data related to forest fire prediction, including meteorological, terrain, vegetation, and human activity data, in March and April 2015–2020. For the mentioned multi-source heterogeneous data, we constructed a diversified knowledge extraction method, which converts the data into triples according to the semantics of the concept layer. It is driven by controller, and its functions include coordinate system conversion, vector data clipping, raster data clipping, raster calculator, conversion from vector data to raster data, conversion from raster data to vector data, slope calculation, aspect calculation, and relative humidity calculation. Figure 11 shows the schematic diagram of the knowledge extraction process of the aspect. First, we used the Arcpy library to convert the coordinate system of the elevation data from GCS_WGS_1984 to WGS_1984_UTM_Zone_48N and cut the elevation data according to the spatial range of the study area. Next, we used the Arcpy library to calculate the aspect according to the elevation and convert the coordinate system of the aspect data from WGS_1984_UTM_Zone_48N to GCS_WGS_1984. The aspect is described in positive degrees from 0 to 360, which is measured clockwise from the north. Then, it decomposes the aspect separately to the east-west and north-south directions. Next, we converted the numerical range of the data from real numbers to integers using the Arcpy library. We provided a trick for reducing the precision loss caused by the data conversion with the Arcpy library. We multiplied the raster values by 10,000 before conversion and carried out the reverse calculations after the conversion. Finally, we saved the final result as a JSON file.

We transformed the model optimization process into spatio-temporal semantic reasoning rules in the case study with the following rule sets, as shown in Figure 12.

- Definition of spatio-temporal semantic rule set 1: Extracting spatio-temporal data from KGFFP to make labeled and unlabeled datasets.
- Definition of spatio-temporal semantic rule set 2: Inputting the data set into the machine learning model for model training or prediction with the support of inference rules.
- Definition of spatio-temporal semantic rule set 3: Calculating the accuracy based on the prediction results and real data and evaluating the accuracy of the prediction model.

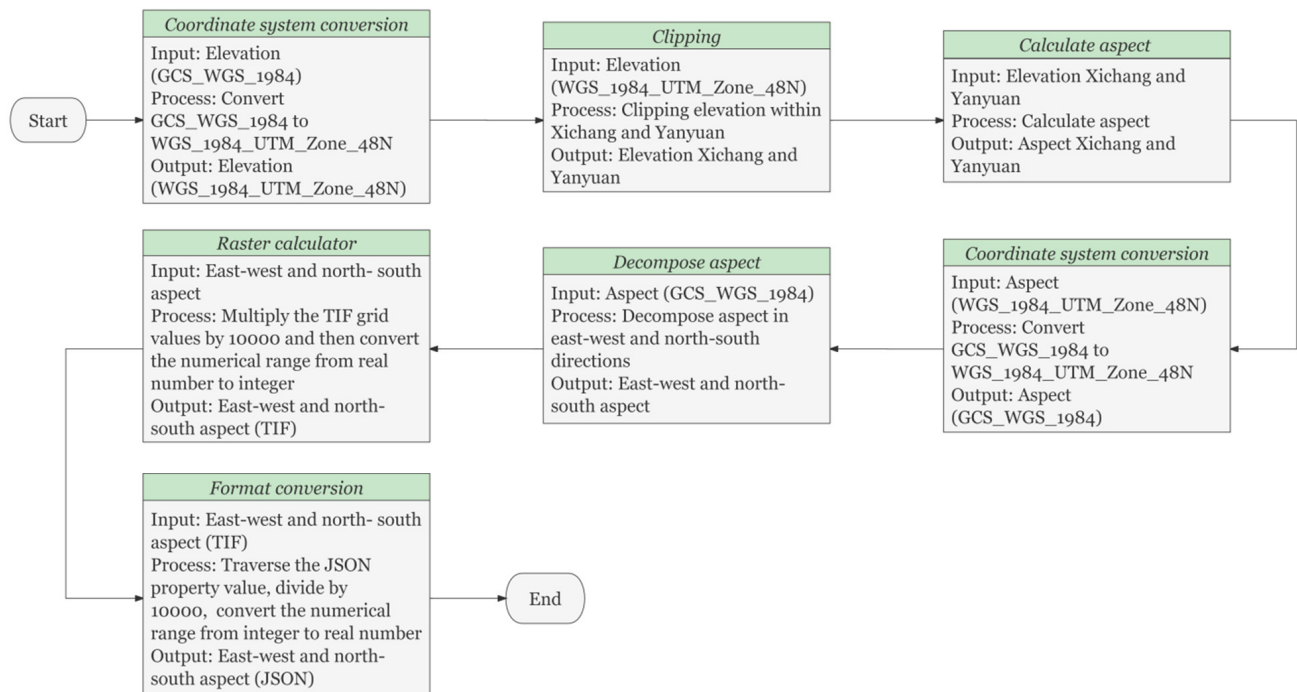


Figure 11. Knowledge extraction process using aspect as an example.

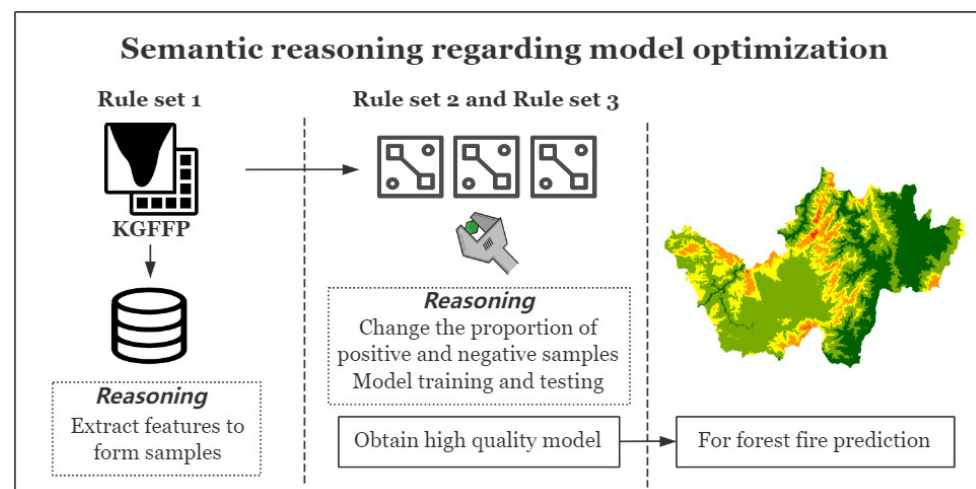


Figure 12. Semantic reasoning regarding model optimization.

3.4. Predicting Forest Fires

This section introduces Sensitivity analysis of the proposed method and the results of the experiment.

3.4.1. Sensitivity Analysis

The optimal ratio of the number of non-fire points to the number of fire points is needed for reducing the susceptibility to probability calculation errors due to unbalanced sample numbers. We performed a sensitivity analysis method to test the ratio on the RF-based and DF-based forest fire models from 1.0 to 2.0 with a step size of 0.1. The performance of the model was evaluated by the metrics *Precision* and *Recall*. The closer the difference between the metrics and 0, the smaller the precision deviation caused by the imbalance in the number of fire samples and non-fire samples.

3.4.2. Experimental Results

1. Experiment 1

Experiment 1 worked to predict the forest fires in Xichang City in March and April 2020 based on the fire data of Xichang City from March and April of 2015 to 2019. Figure 13a,d show a fire map of Xichang in March and April 2015–2019 and 2020 [12].

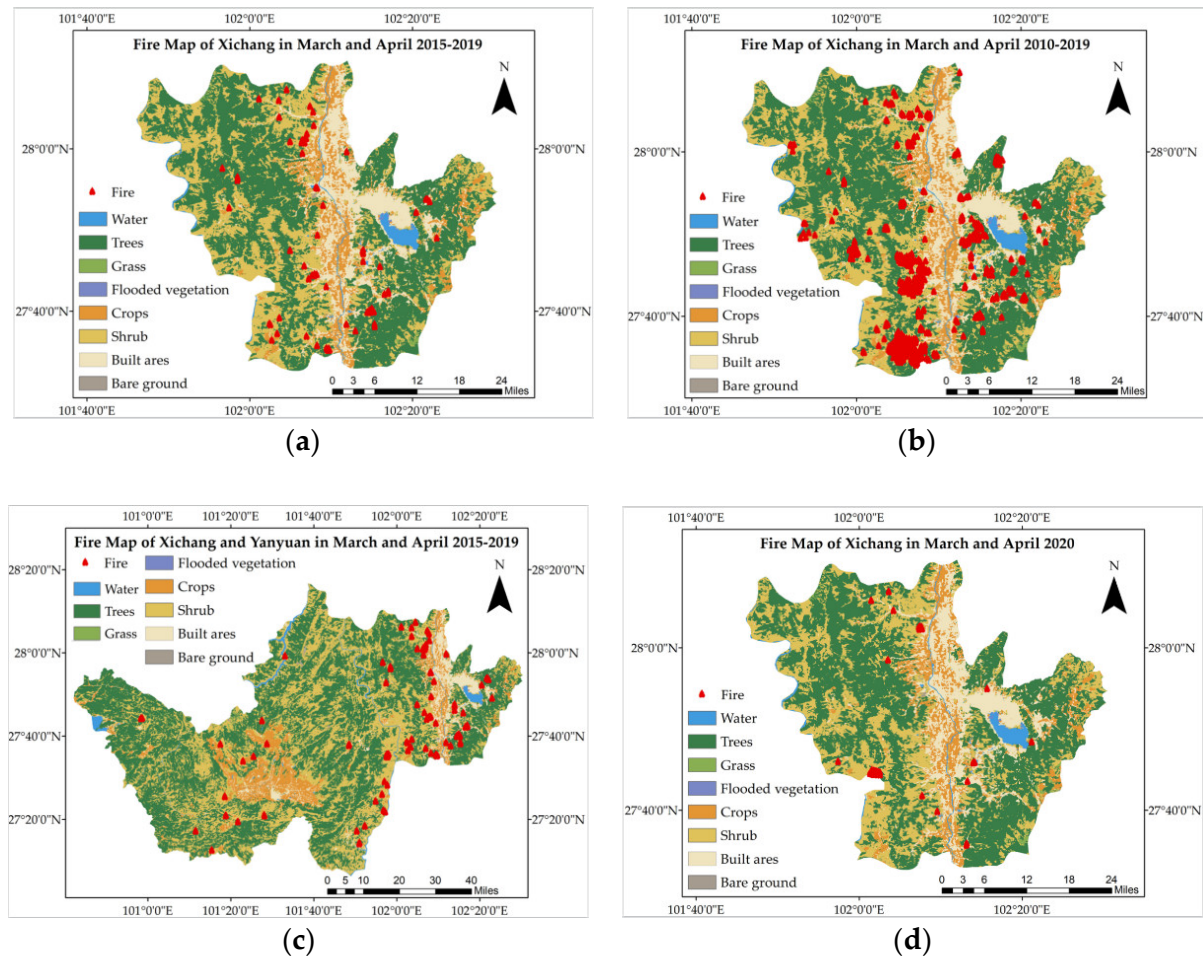


Figure 13. (a) Fire maps of Xichang in March and April 2015–2019. (b) Fire maps of Xichang in March and April 2010–2019. (c) Fire maps of Xichang and Yanyuan in March and April 2015–2019. (d) Fire maps of Xichang in March and April 2020.

Based on rule set 1, we extracted the spatio-temporal knowledge of fire, meteorology, terrain, vegetation, and human factors from the KGFFP in 2015–2019 and 2020. We used this spatio-temporal knowledge to produce training samples and test samples for machine learning-based forest fire prediction models. First, the real fire point was extracted according to the land cover type of the fire point data in March and April 2015–2020. Next, positive samples were made based on real fire points in March and April 2015–2020, and positive samples were constructed based on the fire point data in March and April of 2015–2020 in Xichang City. Then, negative samples were made based on unfired areas and β in March and April 2015–2020, and negative samples were constructed in the unfired area as a training set. After that, the samples from March and April 2015–2019 were set as the training samples and the samples from March and April 2020 were set as the test samples. Finally, training samples and test samples were obtained. The ratio of negative samples to positive samples was $\beta \in [1.0, 2.0]$.

Based on rule set 2, we trained and tested the RF-based and the DF-based forest fire models, respectively, with a step size of 0.1. We then obtained the predicted results of RF-based and DF-based forest fire models.

Based on rule set 3, we compared the actual and predicted values of the Xichang fire point in March and April 2020 and evaluated the predicted results. The meanings of *Precision* and *Recall* are shown in Formulas (1) and (2).

$$Precision = \frac{TP}{(TP + FP)} \quad (1)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (2)$$

where *TP* (True Positive) means that the prediction is correct and the sample is positive; *FP* (False Positive) means that the prediction is wrong and the sample is predicted to be positive, but the sample is actually negative; *FN* (False Negative) means that the prediction is wrong and the sample is predicted to be negative, but the sample is actually positive.

For the RF model, when $\beta = 1.3$, the difference between *Precision* and *Recall* reaches the minimum value. Moreover, the values of *Precision* and *Recall* are relatively high, which is satisfactory. The definition of the metric *F1* is shown in Formula (3). The result of the *F1* of the RF-based forest fire model is 0.7839. For the DF-based forest fire model, when $\beta = 1.5$, the difference between *Precision* and *Recall* reaches the minimum value, and *F1* = 0.7957. Figure 14 shows the experimental results. Table 3 shows the statistical information of samples. Figure 15 shows the probability map of forest fire risk for March and April 2020 for Experiment 1. The prediction results of between the prediction methods with RF and with DF are not drastically different because both are decision tree-based machine learning models. The prediction methods with both models are significantly different from the prediction results with SVM. In our experiment, we select RF and DF for forest fire models because they have relatively high prediction accuracies.

$$F1 = \frac{2}{\frac{1}{P} + \frac{1}{R}} = 2 \times \frac{P \times R}{P + R} \quad (3)$$

where *F1* is the harmonic mean of *Precision* and *Recall*, *P* indicates *Precision*, and *R* indicates *Recall*. The larger the metric *F1*, the better the overall performance of the model. High *F1* value means both *Precision* and *Recall* are good.

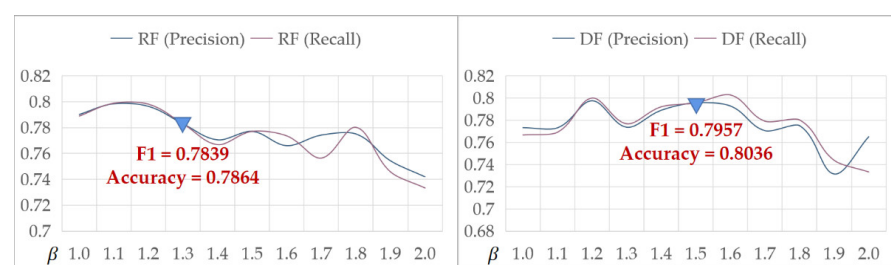


Figure 14. The training set is from March and April of 2015–2019 in Xichang City and the test set is from March and April of 2020 in Xichang City. The symbol “▼” denotes the difference between *Precision* and *Recall* reaches the minimum value.

Table 3. The number of samples, when the training set is from March and April of 2015–2019 in Xichang City, the test set is from March and April of 2020 in Xichang City.

β	Training Data		Test Data	
	Positive Sample	Negative Sample	Positive Sample	Negative Sample
1.3	73	94	45	58
1.5	73	109	45	67

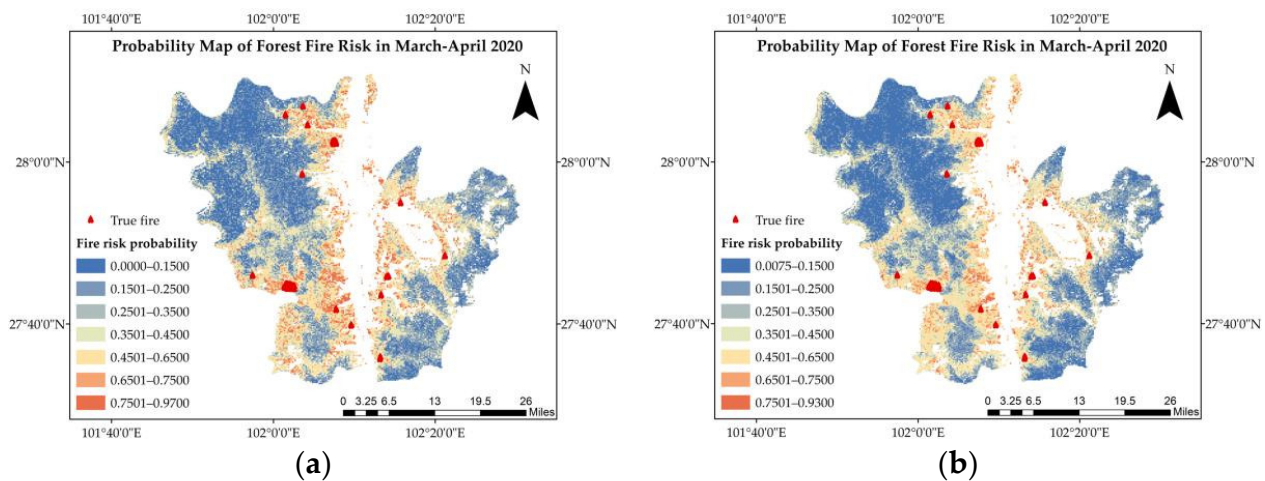


Figure 15. Probability map of forest fire risk in March–April 2020 for Experiment 1 (based on fire data in Xichang from March and April 2015 to March and April 2019). (a) Prediction based on RF. (b) Prediction based on DF.

Based on data used by the above experiment, we added the data of Xichang City from March and April of 2010 to 2014. Based on the fire point data of Xichang City from March and April of 2010 to 2019, it can predict the forest fires of Xichang City in March and April of 2020. Figure 13b,d show the fire map of Xichang in March and April 2010–2019 and 2020 [12].

Positive samples (quantity is 659) are constructed based on the fire point data in March and April of 2010–2019 in Xichang City, and negative samples are constructed in the unfired area as a training set. Positive samples (quantity is 45) are constructed based on the fire data of Yanyuan County in March and April 2020, and negative samples are constructed in the unfired area as a test set. Figure 16 shows the metric *F1* and the accuracy of the RF model and the DF model. Table 4 shows the statistical information of samples. Figure 17 shows the probability map of forest fire risk for March and April 2020 based on the data of Xichang City from March and April of 2010 to 2019. The prediction results between the prediction methods with RF and with DF are less different because both are decision tree-based machine learning models. The prediction methods with both models are significant different from the prediction results with SVM. In our experiment, we select RF and DF for forest fire models because they have relatively high prediction accuracies.

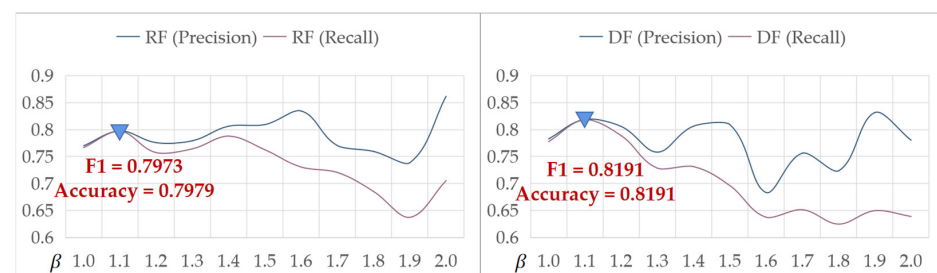


Figure 16. The training set is from March and April of 2010–2019 in Xichang City and the test set is from March and April of 2020 in Xichang City. The symbol “▼” denotes the difference between *Precision* and *Recall* reaches the minimum value.

Table 4. The number of samples, when the training set is from March and April of 2010–2019 in Xichang City, the test set is from March and April of 2020 in Xichang City.

β	Training Data		Test Data	
	Positive Sample	Negative Sample	Positive Sample	Negative Sample
1.1	659	724	45	49

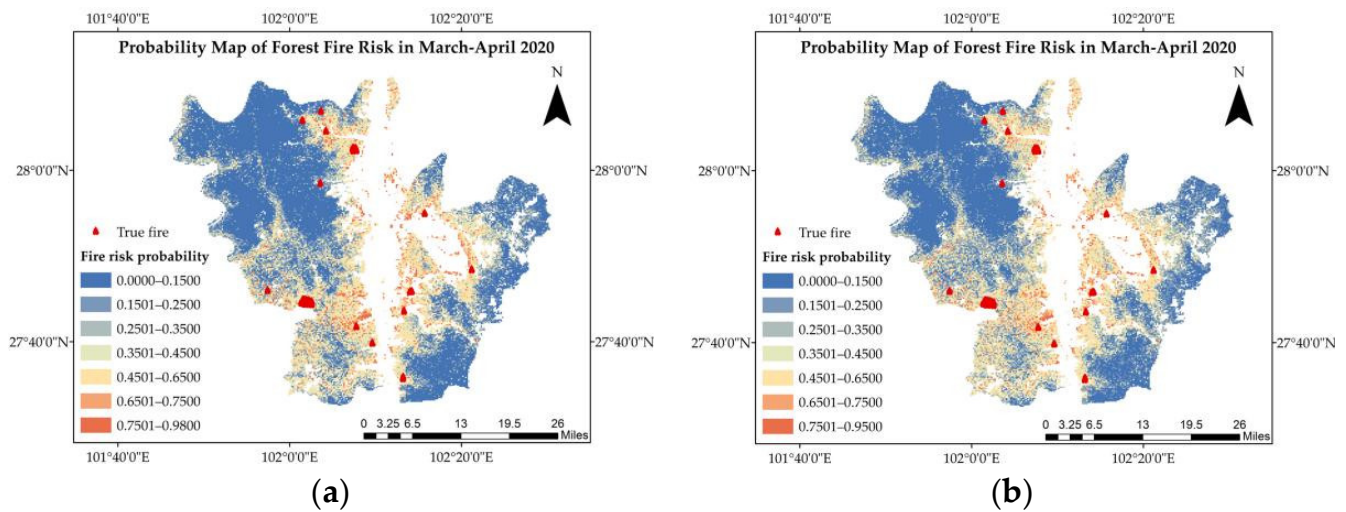


Figure 17. Probability map of forest fire risk in March and April 2020 for Experiment 1 (based on fire data in Xichang from March and April 2010 to March and April 2019). (a) Prediction based on RF. (b) Prediction based on DF.

2. Experiment 2

Since Xichang City and Yanyuan County are adjacent in space, factors such as meteorology, topography, and vegetation are similar, to a certain extent. In that case, the data of Yanyuan County is used as a part of the training samples of Xichang City to predict the accuracy of fire points in Xichang City. Figure 13c,d shows a fire map of Xichang and Yanyuan in March and April 2015–2019 and a fire map of Xichang in March and April 2020 [12].

Positive (quantity is 150) and negative samples were constructed in the unfired area as a training set. Positive samples (quantity is 45) were constructed based on the fire point data in March and April 2020 in Xichang City, and negative samples were constructed in the unfired area as a test set. Figure 18 shows the metric *F1* and accuracy of the RF model and DF model. Table 5 shows the statistical information of samples. Figure 19 shows the probability map of forest fire risk for March–April 2020 for Experiment 2. The prediction results between the prediction methods with RF and with DF are not drastically different because both are decision tree-based machine learning models. The prediction methods with both models are significant different from the prediction results with SVM. In our experiment, we select RF and DF for forest fire models because they have relatively high prediction accuracies.

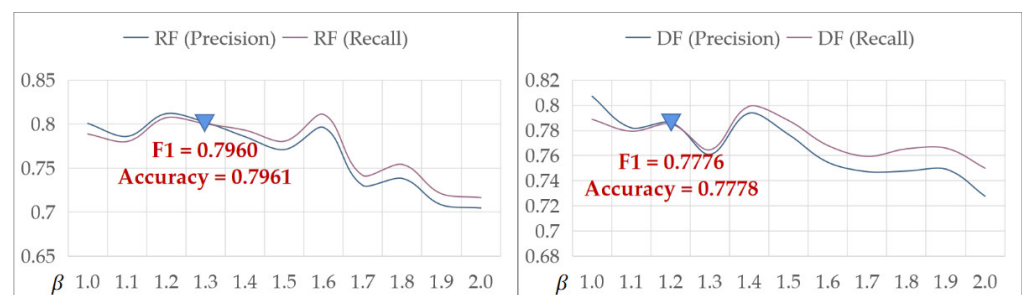


Figure 18. The training set comes from Xichang City and Yanyuan County in March and April 2015–2019 and the test set comes from Xichang City in March and April 2020. The symbol “▼” denotes the difference between *Precision* and *Recall* reaches the minimum value.

Table 5. The number of samples, when the training set is from March and April of 2015–2019 in Xichang City and Yanyuan County, the test set is from March and April of 2020 in Xichang City.

β	Training Data		Test Data	
	Positive Sample	Negative Sample	Positive Sample	Negative Sample
1.2	150	180	45	54
1.3	150	195	45	58

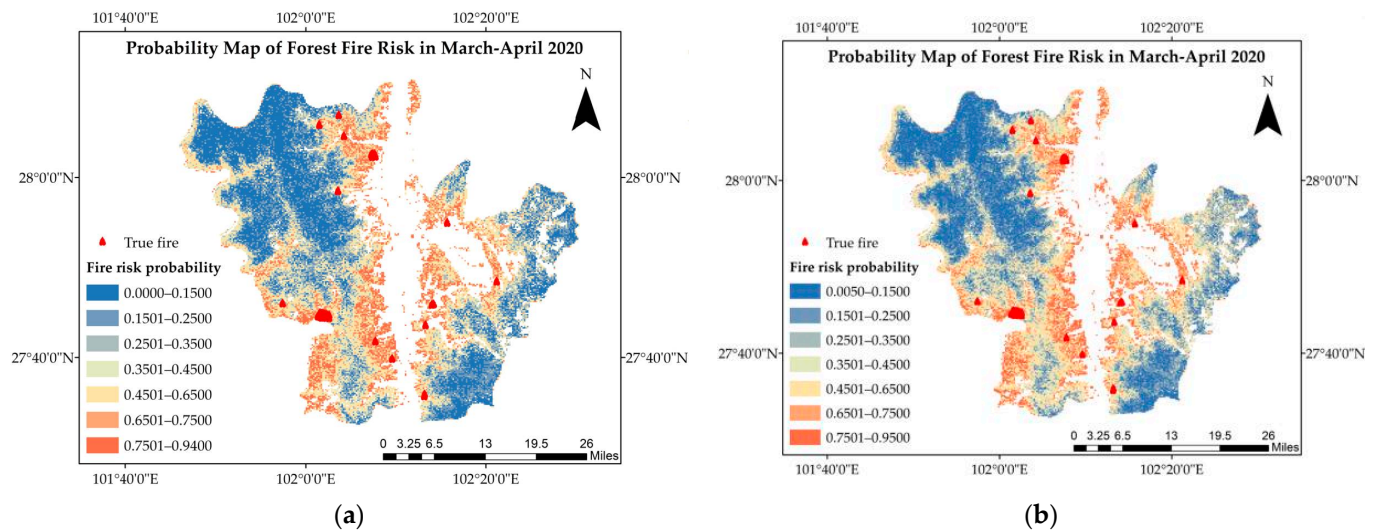


Figure 19. Probability map of forest fire risk in March and April 2020 for Experiment 2. (a) Prediction based on RF. (b) Prediction based on DF.

4. Discussion

Experiment 1 used two datasets for Xichang city. The first was fire data from March and April 2010 to 2014, and the second added the fire data of Xichang from 2015 to 2019. Experiment 2 was based on Experiment 1. The first dataset was used, the test set was unchanged, and the data of Yanyuan County from 2015 to 2019 was added to the training set of Xichang City from 2015 to 2019. Yanyuan County and Xichang City are geographically adjacent to each other.

The experimental results show that the $F1$ of the first dataset is lower than that of the second dataset in Experiment 1. We speculate that the first training data leads to insufficient training of the model. The experimental results show that the $F1$ of the RF-based forest fire model in Experiment 2 increases more than the Experiment 1 with the first dataset, while the $F1$ of the DF model decreases. Table 6 shows the result of these experiments.

Table 6. The result of the two experiments.

Experiment 1				Experiment 2	
The First Dataset		The Second Dataset		$F1$ (RF)	$F1$ (DF)
$F1$ (RF)	$F1$ (DF)	$F1$ (RF)	$F1$ (DF)		
0.7839	0.7957	0.7973	0.8191	0.7960	0.7776

The method proposed in this paper has a certain novelty in multi-source heterogeneous spatio-temporal data fusion and formalization of domain expertise. Our method transforms multi-source heterogeneous spatio-temporal data semantically into spatio-temporal facts in KGFFP and fully preserves their temporal, spatial, and attribute characteristics. Compared with conventional machine learning-based forest fire prediction methods, our method is good at integrating multi-source data for a comprehensive analysis of historical fire points.

It also supports obtaining relatively high-quality analytical datasets in cases where the information is noisy and insufficient. Compared with traditional machine learning-based forest fire prediction, our method has an advantage in that it can formalize domain expertise as inference rules of KGFFP. For forest fire predictions, expertise-oriented data analysis can organize multi-source heterogeneous spatio-temporal data, promote the efficient flow of the prediction analysis process, and obtain high-quality prediction results. Comprehensively considering the above experimental results, we conclude as follows: (1) With the RF model and the DF model, the forest fire prediction method will improve the metric $F1$ when adding different years to the training set in the case that the test set remains unchanged. (2) With the RF model, the forest fire prediction method will improve the metric $F1$ to a certain extent when adding samples from adjacent areas to the training set (samples from Yanyuan County are added to the Xichang training set) in the case that the test set remains unchanged. Therefore, when desiring to improve the metric $F1$ of the RF model, it needs to increase the training samples of the same month and different years in the same area or the samples of the same month and the same year in the adjacent area. (3) With the DF model, the metric $F1$ of the forest fire method cannot be improved when adding samples from adjacent areas to the training set (samples from Yanyuan County are added to the Xichang training set) in the case that the test set remains unchanged. Therefore, when desiring to improve the metric $F1$ of the DF model, it needs to increase the training samples of the same month and different years in the same area.

It is possible to expand our method for future requirements. If we continuously increase the samples of historical March and April to predict forest fires in March and April of 2020, it will lead to a decrease in prediction accuracy. The main reason is that the characteristics of the historical samples and that of 2020 are quite different. This is caused by the passage of time and the great changes in meteorology, topography, vegetation, and human factors. In addition, the experimental results show that it can improve the metric $F1$ of the RF model-based forest fire methods by adding fire points in adjacent regions as training samples, while the similarity of meteorology, topography, vegetation, and human factors between regions needs to be studied. Because the similarity of these factors represents the similarity of training and test samples, it needs a lot of experiments to support improving the metric $F1$.

5. Conclusions

In this paper, we focus on the fusion of multi-source heterogeneous data and the modeling of forest fire expertise. To address the problems, we propose a novel forest fire prediction method for fusing multi-source heterogeneous data in the field of forest fires. It preserves the semantics of describing spatio-temporal facts. We propose a method to formalize domain expertise as semantic rules for knowledge graphs. It helps to obtain the scenario-related data with spatio-temporal features from KGFFP by rule-based reasoning for the machine learning-based forest fire prediction methods.

The main ideas of the proposed method can also be applied to other disaster predictions by suitable extensions. When predicting landslide forecasting using our method, it needs to collect spatio-temporal data closely related to the occurrence of landslides to expand the knowledge graph. Based on the machine learning model concept ontology, our method can support landslide prediction using machine learning models. It needs to choose a suitable machine learning model suitable for landslide prediction. Based on the characteristics of this machine learning model, our method enriches the machine learning model concept ontology and inference rules. The controller triggers the inference rules to invoke the machine learning model for landslide prediction.

However, to clearly describe the proposed method, this paper takes forest fire as an example to introduce the new ideas. It contains mechanism, model, process, and functions of the method that can effectively improve the forest fire predictions. The contributions of this paper are as follows.

- (1) KGFFP integrates multi-source heterogeneous data through semantic technology from the perspective of cross-domain data integration;
- (2) This paper proposes a method to model the domain expertise. It can effectively represent multi-source expertise with a triples form that can facilitate optimization and prediction of the machine learning models for forest fire prediction scenarios;
- (3) Relying on the proposed method, the machine learning-based forest fire prediction methods can be optimized according to historical data with satisfied accuracies. In the case of providing future forest fire-related data, it is expected to obtain better forest fire prediction results.

The occurrence of forest fires is greatly affected by temporal and spatial characteristics. The distribution of geographical data in different regions is very different. Therefore, the applications of forest fire prediction models are often limited to a specific spatial range, and it is difficult to transfer and reuse the method. The proposed method aims to integrate real-world multi-source heterogeneous spatio-temporal data and improve forest fire predictions. In the future, we will focus on domain adaptation in the field of computer vision in the forest fire prediction process. We will build a forest fire prediction model based on domain adaptation and multilayer perceptron. Additionally, to verify the generalization of the method, we will conduct model transfer experiments.

Author Contributions: Conceptualization, X.G., L.P. and Y.Y.; methodology, X.G., L.P. and W.L.; validation, X.G. and W.L.; resources, L.P.; data curation, X.G., W.L., W.Z. and J.C.; writing—original draft preparation, X.G. and L.C.; writing—review and editing, X.G., Y.Y. and L.P.; funding acquisition, L.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Ningxia Key R&D Program (2020BFG02013), and the Beijing Municipal Science and Technology Project (Z191100001419002).

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Podur, J.; Martell, D.L.; Csillag, F. Spatial patterns of lightning-caused forest fires in Ontario, 1976–1998. *Ecol. Model.* **2003**, *164*, 1–20. [CrossRef]
2. Hering, A.S.; Bell, C.L.; Genton, M.G. Modeling spatio-temporal wildfire ignition point patterns. *Environ. Ecol. Stat.* **2008**, *16*, 225–250. [CrossRef]
3. Salavati, G.; Saniei, E.; Ghaderpour, E.; Hassan, Q.K. Wildfire Risk Forecasting Using Weights of Evidence and Statistical Index Models. *Sustainability* **2022**, *14*, 3881. [CrossRef]
4. Chen, J.; Wang, X.; Yu, Y.; Yuan, X.; Quan, X.; Huang, H. Improved Prediction of Forest Fire Risk in Central and Northern China by a Time-Decaying Precipitation Model. *Forests* **2022**, *13*, 480. [CrossRef]
5. Ge, X.; Yang, Y.; Chen, J.; Li, W.; Huang, Z.; Zhang, W.; Peng, L. Disaster Prediction Knowledge Graph Based on Multi-Source Spatio-Temporal Information. *Remote Sens.* **2022**, *14*, 1214. [CrossRef]
6. Jaafari, A.; Gholami, D.M.; Zenner, E.K. A Bayesian modeling of wildfire probability in the Zagros Mountains, Iran. *Ecol. Inform.* **2017**, *39*, 32–44. [CrossRef]
7. Sakr, G.E.; Elhajj, I.H.; Mitri, G.; Wejinya, U.C. Artificial intelligence for forest fire prediction. In Proceedings of the 2010 IEEE/ASME International Conference on Advanced Intelligent Mechatronics 2010, Montreal, QC, Canada, 6–9 July 2010; pp. 1311–1316.
8. PPham, B.T.; Jaafari, A.; Avand, M.; Al-Ansari, N.; Dinh Du, T.; Yen, H.P.H.; Phong, T.V.; Nguyen, D.H.; Le, H.V.; Mafi-Gholami, D.; et al. Performance Evaluation of Machine Learning Methods for Forest Fire Modeling and Prediction. *Symmetry* **2020**, *12*, 1022. [CrossRef]
9. Ma, W.Y.; Feng, Z.K.; Cheng, Z.X.; Wang, F.G. Study on driving factors and distribution pattern of forest fires in Shanxi province. *J. Cent. South Univ. For. Technol.* **2020**, *40*, 57–69.
10. Singh, M.; Huang, Z. Analysis of Forest Fire Dynamics, Distribution and Main Drivers in the Atlantic Forest. *Sustainability* **2022**, *14*, 992. [CrossRef]
11. Prapas, I.; Kondylatos, S.; Papoutsis, I.; Camps-Valls, G.; Ronco, M.; Fernández-Torres, M.; Guillem, M.P.; Carvalhais, N. Deep Learning Methods for Daily Wildfire Danger Forecasting. *arXiv* **2021**, arXiv:2111.02736.
12. Fire Map—NASA. Available online: <https://firms2.modaps.eosdis.nasa.gov/map/> (accessed on 29 April 2022).

13. Cui, L.; Luo, C.; Yao, C.; Zou, Z.; Wu, G.; Li, Q.; Wang, X. The Influence of Climate Change on Forest Fires in Yunnan Province, Southwest China Detected by GRACE Satellites. *Remote Sens.* **2022**, *14*, 712. [CrossRef]
14. Tavakkoli Piralilou, S.; Einali, G.; Ghorbanzadeh, O.; Nachappa, T.G.; Gholamnia, K.; Blaschke, T.; Ghamisi, P. A Google Earth Engine Approach for Wildfire Susceptibility Prediction Fusion with Remote Sensing Data of Different Spatial Resolutions. *Remote Sens.* **2022**, *14*, 672. [CrossRef]
15. Schulte, L.A.; Mladenoff, D.J. Severe wind and fire regimes in northern forests: Historical variability at the regional scale. *Ecology* **2005**, *86*, 431–445. [CrossRef]
16. Ali, A.A.; Carcaillet, C.; Bergeron, Y. Long-term fire frequency variability in the eastern Canadian boreal forest: The influences of climate vs. local factors. *Glob. Chang. Biol.* **2009**, *15*, 1230–1241. [CrossRef]
17. Chen, J.; Ge, X.; Li, W.; Peng, L. Construction of spatio-temporal Knowledge Graph for Emergency Decision Making. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 3920–3923.
18. O'Connor, M.J.; Das, A.K. A method for representing and querying temporal information in owl. In Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies, Valencia, Spain, 20–23 January 2010; pp. 97–110.
19. Chen, J.; Zhong, S.; Ge, X.; Li, W.; Zhu, H.; Peng, L. Spatio-Temporal Knowledge Graph for Meteorological Risk Analysis. In Proceedings of the 2021 IEEE 21st International Conference on Software Quality, Reliability and Security Companion (QRS-C), Hainan, China, 6–10 December 2021; pp. 440–447.
20. What is Arcpy?—Help | ArcGIS for Desktop. Available online: <https://desktop.arcgis.com/en/arcmap/10.3/analyze/arcpy/what-is-arcpy-.htm> (accessed on 14 July 2022).
21. ERA5-Land Hourly Data from 1950 to Present. Available online: <https://cds.climate.copernicus.eu/> (accessed on 8 June 2022).
22. LP DAAC—Homepage. Available online: <https://lpdaac.usgs.gov/> (accessed on 8 June 2022).
23. Esri_2020_Land_Cover_V2 ImageServer. Available online: https://tiledimageservices.arcgis.com/P3ePLMYs2RVChkJx/arcgis/rest/services/Esri_2020_Land_Cover_V2/ImageServer (accessed on 8 June 2022).
24. National Forest Resources Intelligent Management Platform. Available online: <http://www.stgz.org.cn/> (accessed on 8 June 2022).
25. Data Center of Resources and Environment Science of Chinese Academy of Sciences. Available online: <http://www.resdc.cn> (accessed on 6 January 2022).
26. Li, X.H. Using “random forest” for classification and regression. *Chin. J. Appl. Entomol.* **2013**, *50*, 1190–1197.
27. Zhang, L.; Wang, L.L.; Zhang, X.D.; Liu, S.R.; Sun, P.S.; Wang, T.L. The basic principle of random forest and its applications in ecology: A case study of *Pinus yunnanensis*. *Acta Ecol. Sin.* **2014**, *34*, 650–659.
28. Fang, K.N.; Wu, J.B.; Zhu, J.P.; Xie, B.C. A review of random forest method research. *Stat. Inf. Forum* **2011**, *26*, 32–38. (In Chinese)
29. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
30. Lecun, Y.; Boser, B.; Denker, J.S. Handwritten digit recognition with a back-propagation network. In *Advances in Neural Information Processing Systems*; Morgan Kaufmann Publishers: San Francisco, CA, USA, 1990; pp. 396–404.
31. Kotschieder, P.; Fiterau, M.; Criminisi, A. Deep neural decision forests. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1467–1475.
32. Zhen, X.; Wang, Z.; Islam, A.; Bhaduri, M.; Chan, I.; Li, S. Multi-scale deep networks and regression forests for direct biventricular volume estimation. *Med. Image Anal.* **2016**, *30*, 120–129. [CrossRef] [PubMed]